Data Warehousing and Data Mart for Analyzing Arizona and Alabama Health Insurance Rates in 2014

Jeremiah Reyes, Jerry Duran, Chanel Manzanillo
California State University Sacramento
Sacramento, CA

Abstract

Health Insurance Rates can often contain a plethora of information and can be hard to interpret for those who are not exposed to health insurance terminology. Using data from the Center for Medicare and Medicaid Service for the year 2014, we will explore the different rates varying by age as well as examine several different attributes of health care plans for the states Arizona and Alabama by displaying the data in a visual user-friendly manner.

Keywords: Health Insurance Rates, Data Warehouse, Data Mart, Tableau Public

1. Introduction

Health Insurance Rates can often contain a plethora of information and can be hard to interpret for those who are not exposed to health insurance terminology. Using data from the Center for Medicare and Medicaid Service for the year 2014, we will explore the different rates varying by age as well as examine several different attributes of health care plans for the states Arizona and Alabama through data warehousing. We want to analyze Health Insurance plans for Arizona and Alabama in terms of health insurance rates for age and plan, as well as finding other relevant data from the attributes provided. queries will extract data for certain states as well as for certain attributes to show a comparison between the states and different health care plans. We will display the data in queries and a data mart to answer the following questions:
● How do plan rates vary by age?
● How do plan benefits vary across states?
● How do plan types vary across states?

2. Data

The data for the project was obtained through Kaggle [1] which was referenced from the Center for Medicare and Medicaid Service for the years 2014 to 2016. The dataset retrieved contained the cvs files Rate and PlanAttributes. Before data cleaning and data preprocessing, it contained 2 million rows with 24 dimensions. Dimensions include StateCode, Business Year, IssuerId, FederalTIN, PlanId, Age, SourceName,

and RatingAreaId. For both files columns including IssuerID and FederalTIN were taken out since the only data retrieved would be from the year 2014. Columns were taken out that did not have meaning to the average person and simplified to be understood. Null values were removed from the data and data for the years 2015 and 2016 were removed since the data did not contain as many attributes as the year 2014. After data cleaning and data preprocessing, we reduce the data to two csv files Rate and PlanAttributes. Rate now has 216283 rows with 4 columns and PlanAttributes has 657 rows with 9 columns.

3. Design and Methodology

To analyze health Insurance plans for Arizona and Alabama and to answer the data warehousing questions a star schema was constructed.
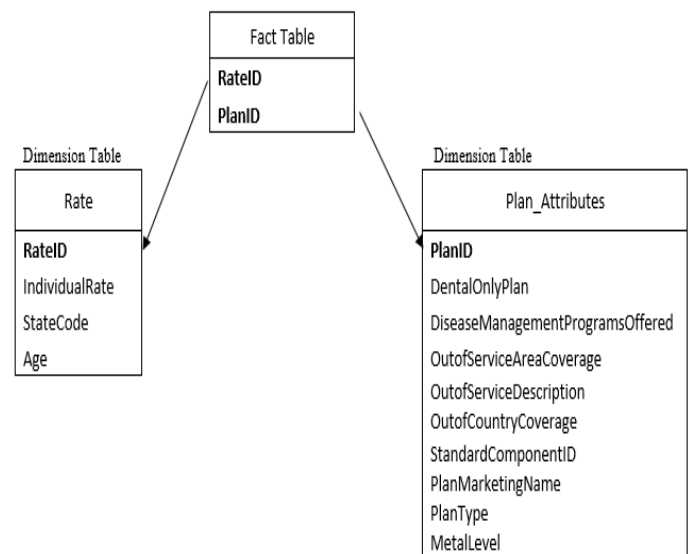

Figure 1. Star Schema

In the star schema (Figure 1) there are two dimensional tables connected with the fact table. The fact table contains the two main primary keys: RateID and PlanID.

The Rate table contains information pertinent to the health insurance rates for the states Alabama and Arizona in the year 2014. Several different attributes are listed in this table such as RateID, which is the unique id classifying the types of health insurance. The attribute IndividualRate which is the cost of the rate an individual pays. The attribute StateCode which is the identification of the State (ie. AZ for Arizona and AL for Alabama). Lastly, the attribute Age which is the age of the person for which the particular health plan is offered to, is a string value from the ages 0-20, to over 65 years old.

The Plan_Attributes table contains information pertinent to the type of ammenities offered through the plans and if a certain healthcare plan covers the listed attributes in this table. The PlanID attribute lists the the unique healthcare plan id. The DentalOnlyPlan attribute shows if a dental plan is offered in the health insurance. The DiseaseManagementProgramsOffered lists the different types of diseases that plans are offered. The OutofServiceAreaCoverage attribute shows if healthcare is provided out of the service area. The OutofServiceDescription attribute provides more information about the OutofServiceAreaCoverage attribute. The StandardComponentID lists the rate id of the plan. The PlanMarketingName lists the name of the health insurance plan. The PlanType lists the type of health insurance plan (ie. Preferred provider organization (PPO)). The OutofCountryCoverage attribute shows if healthcare is provided outside the United States. TheOutofCountryCoverageDescription attribute provides additional details to the OutofCountryCoverage attribute. The MetalLevel attribute shows if a plan is rated Gold, Silver, Bronze.

3.1 Data Warehouse

SQLite Manager [2] was used to construct our data warehouse. The Rate and PlanAttributes cvs files were imported through the SQLite Manager and the corresponding star schema (Figure 1) was used to relate the two data files together.

To answer the question *"How do plan rates vary by age"* The following queries were used using SQL [3] to extract data for both Alabama and Arizona, in which the state, rate of the health insurance, age, and name of the health plan were displayed.
/ select Arizona state plans with rate and age */*
SELECT StateCode, IndividualRate, Age, PlanMarketingName
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AZ";

/ select Alabama state plans with rate and age */*
SELECT StateCode, IndividualRate, Age,
PlanMarketingName
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AL";
Figure 2 shows the corresponding query for the state Arizona in SQLite Manager, while Figure 3 shows the query for Alabama.

To answer the question *"How do plan benefits vary across states?"* The following queries were used to extract data:
/* select Arizona state benefits */
SELECT PlanMarketingName,
DiseaseManagementProgramsOffered,
OutofCountryCoverage, OutofServiceAreaCoverage
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AZ";
/* select Alabama state benefits */
SELECT PlanMarketingName,
DiseaseManagementProgramsOffered,
OutofCountryCoverage, OutofServiceAreaCoverage
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AL";
The query for health plan benefits done in SQLite Manager for Arizona is shown in Figure 4, and the query for health Alabama's plan benefits is shown in Figure 5.

To answer the question *"How do plan types vary across states?"* the following queries were created to extract the data:
/* select Arizona state plan types */
SELECT PlanMarketingName, PlanType
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AZ";
/* select Alabama state plan types */
SELECT PlanMarketingName, PlanType
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AL";
The query for health plan types done in SQLite Manager for Arizona is shown in Figure 6, and the query for health Alabama's plan types is shown in Figure 7.

3.2 Data Mart

Tableau Public [4], a free online data visualization tool, was used to visualize the data and answer to the three questions. The data warehousing files were imported into Tableau and through creating charts in the software's interface we mimicked the queries through the use of filters. A Health Insurance Rate data mart [5] was created and can be accessed at https://public.tableau.com/profile/chanel4820#!/vizhome/Book2_16899/DashboardAttributes.

Figure 8 shows the visual to answer the question

"How do plan rates vary by age?". Figure 9 shows the chart to answer the visual *"How do plan benefits vary across states?"*. Figure 10 shows the visual that answers the question *"How do plan types vary across states?"*. Using the filters on the right-hand side of the chart, indicated by checkboxes, one can select AZ (Arizona) or AL (Alabama) to generated the information for the specific states and to show the queries for Arizona and Alabama visually.

4. Results

Through analyzing the query generated values and through using filters on the data mart charts the following conclusions were made for the results. Answering the question *"How do plan rates vary by age?"* as age increases, the rate of the health plan also increases. This was found true in both Arizona and Alabama. Answering the question *"How do plan benefits vary across states?"*, both Arizona and Alabama offer out of service and out of country coverage in their various health plans. Arizona offers more plans for "diabetes" and "diabetes and asthma" plans. Alabama offers more plans for "heart disease, diabetes, and pregnancy", and offers no plans to solely cover "diabetes". Answering the question *"How do plan types vary across states?"*, in Alabama HMO (health maintenance organization) plan types are not available. In Arizona POS (point of service) plan types are not available, but have more PPO (preferred provider organization) plans than Arizona.

5. Summary and Conclusions

Using SQLite and Tableau Public, we were able to visualize and extract data relevant in answering questions about the Health Insurance Rate Plans for the year 2014 in Arizona and Alabama. Through analyzing various iEEE papers available on the ACM website [6,7,8] we could see the layout they used to express their data findings. They [6] also extracted certain attributes to analyze their data and to display it in a meaningful fashion for health insurance. in

We learned how to use SQLite manger and Tableau Public, and how to import cvs files to create meaningful queries and charts. We also learned more about health insurance terms such as the HMO, plan. Tableau is a great visualization tool that is easy to use and you are able to import a created database into the program to generate graphs. The project also gave us a perspective that as we get older, the rates for insurance will be higher so we should get in a habit of good health practices. In "Understanding Health Care Insurance Data through Graph Analytics" [7] they showed how to visualize certain aspects of health insurance data through graphs which helped give an example of how to display our own data.

6. Team Responsiblities

Chanel Manzanillo was in charge of the data warehousing, data mart, and data warehousing and data mart report. Gerardo Duran was in charge of helping out with the data mart and the data mart section of the report. Jeremiah Reyes was in charge of data mining and the data mining report.

7. References

[1] "Health Insurance Marketplace | Kaggle". Kaggle.com. N.p., 2017. Web. 5 May 2017

[2] Nguyen, Dan. "Getting Started With Sqlite Browser | Public Affairs Data Journalism At Stanford University". Public Affairs Data Journalism at Stanford University. N.p., 2017. Web. 5 May 2017.

[3] Dube, Ryan. "Learn SQL Or Create A Simple Database With Sqlite Database Browser". MakeUseOf. N.p., 2017. Web. 4 May 2017.

[4] "Resources". Tableau Public. N.p., 2017. Web. 7 May 2017.

[5] "Health Insurance Rates". Tableau Public. N.p., 2017. Web. 17 May 2017.

[6] Ying-Kai Chen and Yu-Hui Tao. 2014. Analyzing the Healthcare Expenditure of National Health Insurance: A HLM Model Construction on the National Health Insurance Database. In Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia (MoMM '14). ACM, New York, NY, USA, 353-355.

[7] Luis G. Moyano, Ana Paula Appel, Vagner F. de Santana, Marcia Ito, and Thiago D. dos Santos. 2016. GraPhys: Understanding Health Care Insurance Data through Graph Analytics. In Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 227-230.

[8] Mohit Kumar, Rayid Ghani, and Zhu-Song Mei. 2010. Data mining to predict and prevent errors in health insurance claims processing. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10). ACM, New York, NY, USA, 65-74.

8. Appendix:

Figure 2. Arizona Health Rates and Plans for Ages 0-20 to 65 and older

Enter SQL

```
/* select Arizona state plans */

SELECT StateCode, IndividualRate, Age, PlanMarketingName
```

Run SQL   Actions ▾   Last Error: not an error

StateCode	IndividualRate	A
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	0-
AZ	49.02	2
AZ	49.02	2
AZ	49.02	2
AZ	49.02	2
AZ	49.02	2
AZ	49.02	2

Enter SQL

```
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode =
```

Run SQL   Actions ▾   Last Error: not an error

StateCode	IndividualRate	Age	PlanMar
AL	125.26	0-20	Humana
AL	197.28	21	Humana
AL	197.28	22	Humana
AL	197.28	23	Humana
AL	197.28	24	Humana
AL	198.07	25	Humana
AL	202	26	Humana
AL	206.74	27	Humana
AL	214.44	28	Humana
AL	220.75	29	Humana
AL	223.9	30	Humana
AL	228.64	31	Humana
AL	233.38	32	Humana
AL	236.32	33	Humana

Figure 3. Alabama Health Rates and Plans for Ages 0-20 to 65 and older

Enter SQL   Select │ Data M

```
SELECT PlanMarketingName, DiseaseManagementProgramsOffered, OutofCountryCoverage, OutofS
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AZ";
```

Run SQL   Actions ▾   Last Error: not an error

PlanMarketingName	DiseaseManagementProgramsOffered	OutOfCountryCo
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes
CopayComplete Select (Maricopa) 40	Asthma, Diabetes	Yes

Figure 4. Arizona health plan benefits

Enter SQL   Selec

```
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "AL";
```

Run SQL   Actions ▾   Last Error: not an error

PlanMarketingName	DiseaseManagementProgramsOffered	OutO
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes
Blue Choice Platinum for Business	Asthma, Heart Disease, Diabetes, Preg...	Yes

Figure 5. Alabama health plan benefits

Enter SQL

```
SELECT PlanMarketingName, PlanType
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode = "
```

Run SQL   Actions ▾   Last Error: | not an error

PlanMarketingName	PlanType
Health Choice Essential Silver	HMO
Health Choice Essential Silver	HMO
Health Choice Essential Silver	HMO
Health Choice Essential Silver	HMO
Health Choice Essential Silver	HMO
myCigna Health Flex 1900	PPO
myCigna Health Flex 1900	PPO
myCigna Health Flex 1900	PPO
myCigna Health Flex 1900	PPO
myCigna Health Flex 1900	PPO

Figure 6. Arizona health plan types

Enter SQL

```
SELECT PlanMarketingName, PlanType
FROM Rate, PlanAttributes
WHERE RateID = StandardComponentID and StateCode =
```

Run SQL   Actions ▾   Last Error: | not an error

PlanMarketingName	PlanType
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS
UnitedHealthcare Gold HSA (UG-9 with A8)	POS

Figure 7. Alabama health plan types

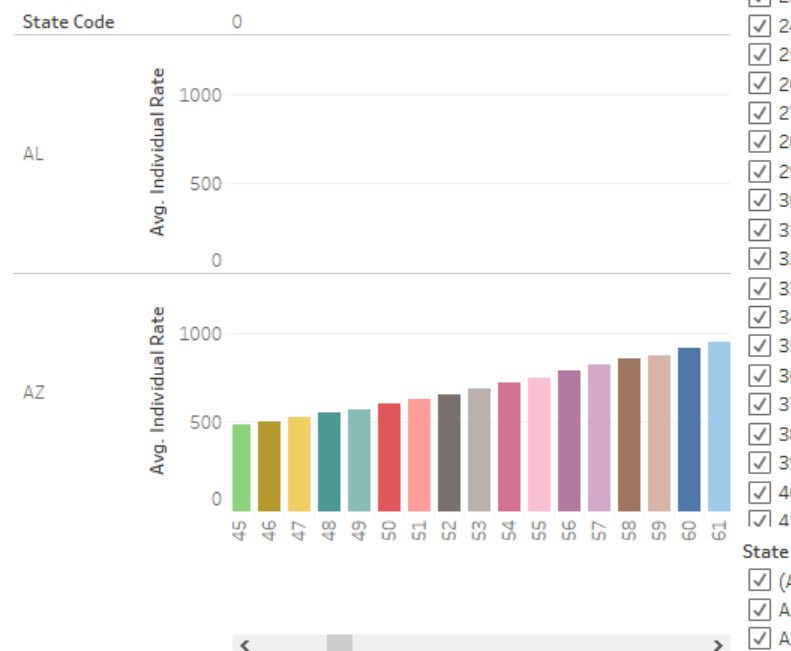**Figure 8. Comparison of Cost and Rate of Alabama and Arizona**

Figure 9. Comparison of Benefits of Alabama and Arizona

Plan Types for Arizona and Alaba

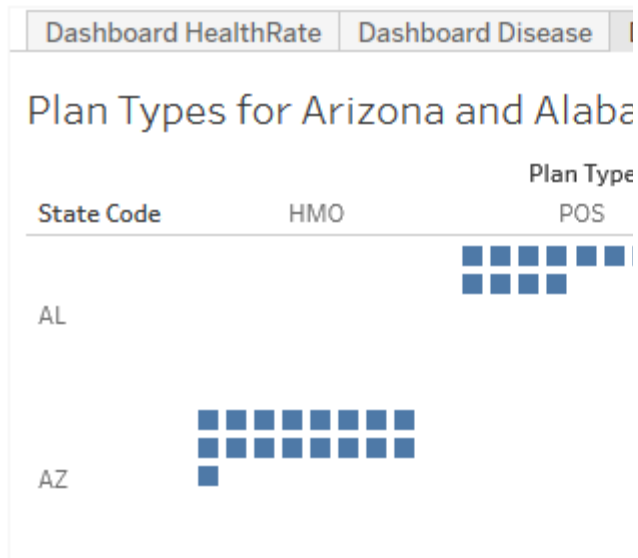|	Plan Type	
State Code	HMO	POS
AL		
AZ		

Figure 10. Comparison of Plan Types of Alabama and Arizona